Medical hypothesis discovery and innovation in ophthalmology

Open Access

Review Article

Application and interpretation of linear-regression analysis

Narges Roustaei [1]

[1] *Ophthalmology Department, IVORC Academic Foundation, Texas, USA*

ABSTRACT

Background: Linear-regression analysis is a well-known statistical technique that serves as a basis for understanding the relationships between variables. Its simplicity and interpretability render it the preferred choice in healthcare research, including vision science, as it enables researchers and practitioners to model and predict outcomes effectively. This article presents the fundamentals of linear-regression modeling and reviews the applications and interpretations of the main linear-regression analysis.

Methods: The primary objective of linear regression is to fit a linear equation to observed data, thus allowing one to predict and interpret the effects of predictor variables. A simple linear regression involves a single independent variable, whereas multiple linear regression includes multiple predictors. A linear-regression model is used to identify the general underlying pattern connecting independent and dependent variables, prove the relationship between these variables, and predict the dependent variables for a specified value of the independent variables. This review demonstrates the appropriate interpretation of linear-regression results using examples from publications in the field of vision science.

Results: Simple and multiple linear regressions are performed, with emphasis on the correct interpretation of standardized and unstandardized regression coefficients, the coefficient of determination, the method for variable selection, assumptions in linear regression, dummy variables, and sample size, along with common mistakes in reporting linear-regression analysis. Finally, a checklist is presented to the editors and peer reviewers for a systematic assessment of submissions that used linear-regression models.

Conclusions: Medical practitioners and researchers should acquire basic knowledge of linear-regression such that they can contribute meaningfully to the development of technology by accurately interpreting research outcomes. Incorrect use or interpretation of appropriate linear-regression models may result in inaccurate results. Appointing an expert statistician in an interdisciplinary research team may offer added value to the study design by preventing overstated results.

KEYWORDS

regression analyses, linear regression, statistics, ophthalmology, sample size, optometry

Correspondence: Narges Roustaei, Ophthalmology Department, IVORC Academic Foundation, Texas, USA.
Email: rousta.biostatistics@gmail.com. ORCID iD: https://orcid.org/0000-0002-5753-2142

How to cite this article: Roustaei N. Application and interpretation of linear-regression analysis. Med Hypothesis Discov Innov Ophthalmol. 2024 Fall; 13(3): 151-159. https://doi.org/10.51329/mehdiophthal1506

Received: 25 June 2024; Accepted: 01 October 2024

INTRODUCTION

Cox regression, logistic regression, Poisson regression, and linear regression are among the different types of regression analysis [1]. Simultaneously analyzing many variables is desirable when predictor factors are associated with an outcome [2]. Regression analysis is typically performing to determine the relationship between one variable and a set of other variables [2, 3]. Statisticians typically denote one variable using Y and refer to it as a dependent, response, outcome, or predicted variable. The other variables are generally denoted by X and referred to as independent variables, explanatory variables, predictor variables, or covariates [2, 3]. The independent variables are numerical, dichotomous, or multicategorical, whereas the dependent variable is numerical [3, 4].

However, if the dependent variable is yes/no or dichotomous, then logistic regression analysis is performed [5, 6]. The presence or absence of computer-vision syndrome is an example of a dichotomous dependent variable [7, 8]. Statistical software packages are used to conduct linear-regression analysis, including statistical packages for the social sciences, as well as statistical-analysis software such as Minitab, R, and Stata [9].

Herein, the application and interpretation of regression analysis as a method for examining variables simultaneously are discussed based on examples pertaining to vision sciences obtained from the literature. The aim is to provide an overview of the components of linear-regression analyses. Subsequently, the applications and interpretations of this field of vision research are explained. Table 1 provides the definitions for the terms used in this review.

Table 1. Terms used in regression analysis

Term	Explanation
Dependent variable	The dependent variable is known as the response, outcome, or predicted variable, which is denoted as "Y." The regression model is used to explain or predict this variable [3, 5].
Independent variable	The independent variable is known as the explanatory variable, predictor variable, or covariates. These variables in the regression model are used to explain or predict changes in the dependent variable and are typically denoted as X [5].
Simple regression analysis	Simple linear regression includes two variables, i.e., a dependent variable (Y) and an independent variable (X), which are linearly related to each other [6].
Multiple-regression analysis	Multiple linear regression includes a dependent variable (Y) and multiple independent variables (X_k), which are linearly related to each other [4].
Regression coefficient	The unstandardized regression coefficient is the amount of change in the dependent variable for a one-unit increase in the continuous independent variable (X_1), whereas other independent variables are fixed. The standardized regression coefficients reveal the independent variables that have the strongest linear relationships with the dependent variable [3, 4].
Coefficient of determination	The coefficient of determination is the proportion of the variation in the dependent variable explained by the regression model [10].
Assumptions in linear regression	Certain assumptions are introduced in linear-regression analysis, and the key ones are normality, equality of variance, linearity, and collinearity [4].

METHODS

Simple linear regression

Simple linear regression is the simplest application of regression involving two linearly related variables, a single dependent variable (Y), and a single independent variable (X), which are linearly related to each other [3, 5, 6]. The objectives of regression analysis are to test for a general underlying pattern connecting two variables and to show the relationship between X and Y to predict Y for a specified value of X [3, 5, 11].

For example, in a study of two series of patients undergoing cataract surgery, an optical biometry group (n = 51 patients) and an ultrasound biometry group (n = 79 patients) were recruited [12]. Four intraocular lens (IOL) power-calculation formulas (Haigis, Hoffer Q, Holladay 1, and SRK/T) were used in both groups. The corneal surface asphericity (Q-value) was obtained by two different Placido-disk corneal topographers in each group. The presence of a relationship between the arithmetic error in refraction prediction (the dependent variable defined as the difference between the expected refraction and post-operative refraction at one month based on the actual IOL implantation according to each power calculation formula) and the Q-value (a single independent variable) was tested using simple linear-regression analysis. The coefficient of determination (R-square [R^2]) was used to determine the proportion of variation in the refraction prediction error explained by the regression model. The refraction prediction error was negative for eyes with post-operative hyperopic refraction and positive for eyes with post-operative myopic refraction [12].

In another study on 115 eyes of 115 patients to assess the relationship between the refraction prediction error (the difference between the expected refraction and the refraction at one-month post operation) and the Q-value, a simple linear regression was performed [13]. Q-values were obtained using three instruments: 115, 104, and 73 eyes were measured using a Scheimpflug–Placido corneal topographer, rotating Scheimpflug camera, and Placido-disk corneal topographer, respectively. Four IOL power-calculation formulas (Haigis, Hoffer Q, Holladay 1, and SRK/T) were used [13]. A scatter diagram with a line drawn through the points was used to summarize the relationship between the two variables or a formula was used to calculate the refraction prediction error based on a specified Q-value [13]. Additionally, the strength and direction of the relationships between the variables can be interpreted. The research objective was to determine the effect of the Q-value on the refraction prediction error. In the scatter diagram illustrated in this study, the Hoffer Q prediction error was measured along the vertical axis, and the Q-value was measured by a Placido-disk corneal topographer along the horizontal axis, with R^2 equal to 0.2630 [13].

The equation for any straight line, except for a vertical line, can be presented in the following form [10]: $\widehat{Y} = a + bX$. Here, "$\widehat{Y}$" denotes ordinate of any point on the line, "a" denotes the Y intercept, and "b" denotes the slope of the line. Meanwhile, "a" and "b" are unstandardized regression coefficients, where "a" is the ordinate or height of the line at point X = 0 (if X = 0 is interpretable). The interpretation of the slope (b) is that when X changes by one unit, the height of the line changes by "b" units [3].

The least-squares method is the most commonly used method for fitting a straight line through a set of points. In the first Q-value example [12], the highest R^2 was indicated between the Hoffer Q formula and the Q-value in the optical biometry group ($R^2 = 0.3050$), for which the refraction prediction error (Y) is related to the Q-value (X) based on the formula; furthermore, this relationship was statistically significant ($P < 0.05$) [12]. In the second Q-value example [13], using the least-squares method, a = 0.2641 and b = - 1.4589 were obtained for the Q-value measured by the Placido-disc corneal topographer and IOL power using the Hoffer Q formula. In this study, the highest R^2 was indicated between the Hoffer Q formula and the Placido-disk corneal topographer ($R^2 = 0.2630$), for which the refraction prediction error (Y) is related to the Q-value (X) based on the formula; furthermore, this relationship was statistically significant ($P < 0.05$) [13]. The regression line computed from a sample is known as the sample regression line, where using the least-squares method, we obtain $\widehat{Y} = 0.2641 + (-1.4589) \times X$. These unstandardized regression coefficients can be used to estimate the refraction prediction error at a specified Q-value. For example, if the Q-value (X) measured by the Placido-disk corneal topographer equals to - 0.2, then the refraction prediction error using the Hoffer Q formula can be calculated as follows: $\widehat{Y} = 0.2641 + (-1.4589)(-0.2) = +0.55588$. A P-value < 0.05 indicates that the independent variable (Q-value) is significantly related to the dependent variable (the refraction prediction error) [13].

Multiple linear regression

The objectives and procedures of simple linear regression can be extended to more than two variables using multiple linear regression. In a multiple-regression model, a single dependent variable (Y) and multiple variables ($X_1, X_2, \ldots, X_K$) are considered to investigate the manner by which these variables in combination are related to Y or can be used to predict Y based on the observed values of $X_1, X_2, \ldots, X_K$ [3, 6, 11, 14].

In another study, Cha et al. [15] included 393 eyes of 215 patients who underwent laser keratorefractive surgery. Among these eyes, laser-assisted in-situ keratomileusis (LASIK), photorefractive keratectomy, and small-incision lenticule extraction was performed on 164 (n = 91 patients), 183 (n = 100 patients), and 46 (n = 25 patients). The study aimed to predict the 1-year post-operative ratio of the posterior-to-anterior curvature radii of the cornea (P/A) using the preoperative characteristic values of refractive surgery patients [15]. For patients who underwent LASIK at a 4 mm zone around the corneal vertex, the relationship between post-operative P/A (Y or dependent variable) and mean manifest refraction spherical equivalent (MRSE) (X_1), preoperative P/A (X_2), posterior mean keratometry (Km) (X_3), age (X_4), and anterior Q-value value (X_5) (X_1, X_2, X_3, X_4, and X_5 or independent variable) was assessed using multiple linear regression [15]. The sample multiple-regression line using the least-squares method was $\hat{Y} = -0.0304 + 0.0154X_1 + 0.9048X_2 - 0.0156X_3 + 0.0004X_4 + 0.0103X_5$. Each of the independent variables in this example exhibited a statistically significant relationship with the dependent variable (all $P < 0.05$) [15].

In this study [15], the unstandardized regression coefficient for the MRSE (X_1) was + 0.0154. This implies that the post-operative P/A increases by 0.0154 for a one-unit increase in the MRSE when the other variables (X_2, X_3, X_4, and X_5) are fixed. Similarly, the unstandardized regression coefficient of - 0.0156 for the posterior Km (X_3) indicates that the post-operative P/A decreases by 0.0156 for a one-unit increase in the posterior Km when the other variables (X_1, X_2, X_4, and X_5) are fixed. The interpretations of the other coefficients reported in this study [15] are the same. All P-values were < 0.05, thus indicating a statistically significant relationship between the post-operative P/A and each independent variable [15]. In this example, the value of - 0.0304 [15] as the intercept is not of particular interest, as values of 0 for each of the independent variables (X_1, X_2, X_3, X_4, or X_5) are impossible; therefore, it does not have a specific interpretation.

In another study, Zareei et al. [16] included 95 eyes of 95 patients, i.e., 47 with primary congenital glaucoma and 48 normal control eyes. They investigated the effects of corneal hysteresis (CH), corneal resistance factor (CRF), and central corneal thickness (CCT) (independent variables) on the Goldman applanation tonometry (GIOP), Goldman intraocular pressure (IOPg), and corneal compensated intraocular pressure (IOPcc) (dependent variables) using multiple-linear-regression analysis. The two groups were matched based on age and gender. They measured the CH, CRF, IOPg, and IOPcc using an ocular-response analyzer; the GIOP using a Goldman applanation tonometer; and the CCT using ultrasonic pachymetry [16]. The multiple-linear-regression equation for patients with PCG and IOPg (the dependent variable) obtained using the least-squares method was $\hat{Y} = \text{intercept} - 1.17\ \text{CH} + 1.45\ \text{CRF} - 0.07\ \text{CCT}$, with all P-values < 0.05 [16], thus indicating that CH, CRF, and CCT were significantly related to IOPg. In this study, the unstandardized regression coefficient for CH was -1.17 [16], which implies that IOPg decreases by 1.7 for a one-unit increase in the CH when other variables, such as the CRF and CCT, are fixed.

Standardized regression coefficients

The size of the unstandardized regression coefficients depends on the units of the measured variables; thus, they are not comparable. For standardized regression coefficients, which are the standardized Y variable regressed on the standardized X variables, the intercept is zero; therefore, the regression coefficients are unitless and directly comparable. By comparing the magnitudes of standardized regression coefficients, researchers can determine the independent variables that have the strongest linear relationships with the dependent variable when other variables are adjusted [3, 4].

In another study, 20 patients with non-proliferative diabetic retinopathy, 20 patients with proliferative diabetic retinopathy, and 20 healthy age-, gender-matched normal controls were recruited; additionally, the mean age, refraction, macular thickness, intraocular pressure, and sex ratio of the study groups were comparable. The aim was to investigate the relationship between the best corrected visual acuity (BCVA) (dependent variable) and optical coherence tomography angiography (OCTA) parameters such as parafoveal and whole area superficial capillary plexus vessel density (SCP-VD), as well as, parafoveal and whole area deep capillary plexus vessel density (DCP-VD) (independent variables), using multiple-linear-regression analysis. The SCP-VD in the entire area and the DCP-VD in the parafoveal area indicated greater standardized regression coefficients compared with the other independent variables, i.e., – 1.38 and – 0.485, respectively (both $P < 0.05$) [17]. The predictability strength for the SCP-VD was determined using stepwise multiple-linear-regression analysis, which indicated that SCP-VD was the most significant predictor of BCVA [17]. Similar to this example [17], standardized regression coefficients should be reported together with unstandardized regression coefficients [17, 18]; both coefficients are available in the output of statistical software packages.

Coefficient of determination

A typical method of assessing the fit of a model is to calculate the R^2, which is based on the output of statistical software. R^2, whose range is 0–1, is used to express the proportion of variation in the dependent variable, as explained by the regression model. When R^2 is approximately 1, most of the variation in Y can be explained by its linear relationship with X. Meanwhile, an R^2 of approximately 0 indicates that the variables are not strongly related or have a relationship other than a linear one, e.g., a quadratic relationship [10]. When additional independent variables are added to the model, the R^2 increases automatically (adjusted for the number of independent variables included in the model). This adjusted coefficient of determination is denoted as R^2_{adj} and adjusts the value of R^2 to account for the number of independent variables included in the model to avoid overestimating the effect of added independent variables to the model [3, 10].

For example, a simple linear-regression analysis for the refraction prediction error using the Hoffer Q IOL power-calculation formula and the Q-value obtained by the Placido-disk corneal topographer, $R^2 = 0.2630$, implies that 26.30% of the variation in the refraction-prediction error is due to its linear relationship with the Q-value, whereas the remaining 73.70% is due to other potential factors [13].

In an example of the multiple linear regression for P/A and the preoperative characteristics of refractive surgery patients after LASIK surgery, an adjusted R^2 of 0.944 for a 4 mm zone around the corneal vertex implies that 94.4% of the variation in the post-operative P/A in eyes that underwent LASIK surgery is due to its linear relationship with the MRSE (X_1), preoperative P/A (X_2), posterior Km (X_3), age (X_4), and anterior Q-value (X_5) with adjustment to the number of independent variables, whereas the remaining 0.6% is due to other potential factors [15]. The high R^2 value in this example [15] indicates that the model with five variables (X_1, X_2, X_3, X_4, and X_5) provides a better fit than the model with only the intercept term.

Method for variable selection

Various methods are used to enter independent variables into the regression model to identify a better combination of variables. Enter, forward, backward, and stepwise selection are among the most common methods. In the "enter method," all independent variables are entered into the regression model at once. The goal of this method is to determine the effect of each independent variable (X) on a dependent variable (Y) [4, 19].

In forward, backward, and stepwise selection methods, independent variables are added or removed in several stages until the remaining variables contribute to the regression fit. Investigators use these methods when numerous independent variables render the interpretability of the final model challenging, and they tend to report an optimal model with the least-significant independent variables [4]. For example, in a study investigating the correlations between OCTA parameters and the BCVA of patients with diabetic retinopathy, stepwise multiple-linear-regression analysis was performed [17].

Assumptions in linear-regression analysis

In simple linear regression, observing a scatter plot allows one to identify the suitability of the model for representing the relationship between two variables [10]. In multiple-regression analysis, assumptions are introduced based on residuals [10] (the difference between the predicted Y [$\hat{Y}$] and observed Y, i.e., $e_i = Y_i - \hat{Y}_i$).

1. **Normality assumptions**

Normality can be assessed from stem and leaf or histogram plots of residuals; if the plot is bell shaped, then the normality assumption is satisfied. Additionally, if the normal quantile plot of the residuals is a straight line, then the normality assumption is satisfied. Moreover, the Shapiro–Wilk statistic can assume values between 0 and 1, and the null hypothesis for this test is that the residual is from a normally distributed population; therefore, $P > 0.05$ indicates that normality is satisfied [4, 10]. In studies with large sample sizes, i.e., the number of observations per variable exceeds 10, violations of the normality assumption do not affect the outcomes [20].

2. **Equality of variance assumptions**

The assumption of equality of variances is assessed using a plot of residuals versus independent variables or a plot of residuals versus fitted values. If the assumption is satisfied, then an apparent pattern should not appear, i.e., a random pattern will be observed and the variability of the residuals will be constant [4, 10] (Figure 1).

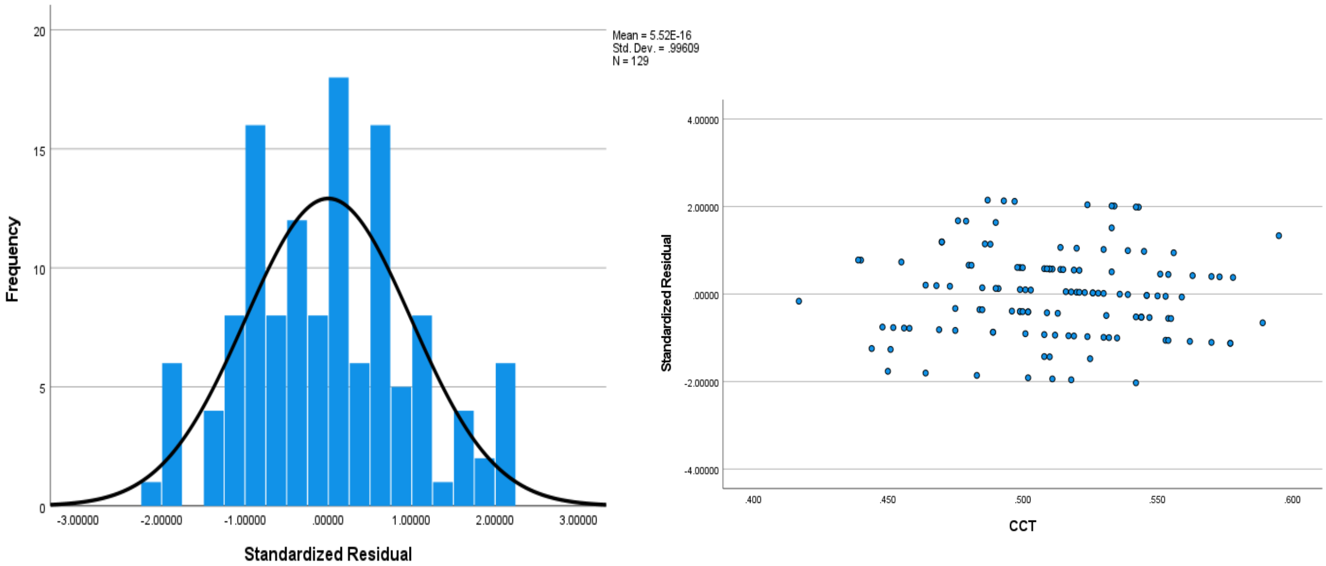

Figure 1. Fictitious example dataset: (*left*) Histogram of distribution of residuals of linear regression for investigating intraocular pressure (IOP) and central corneal thickness (CCT) relationship; (*right*) scatter plot of residuals and CCT values.

3. Linearity assumptions

For k independent variables, k plots of residuals or standardized residuals versus independent variables were drawn; if linearity exists between Y and X_K, then be no apparent pattern should appear in the points. In other words, the slope of the line drawn through the points should be approximately zero [10, 19].

4. Collinearity assumptions

Collinearity refers to the correlation between independent variables when many independent variables are included in a regression model. The estimation of standard errors is challenging when the independent variables are highly correlated. For assessing collinearity, the variance inflation factor (VIF) or tolerance (inverse of the VIF) can be computed [4, 21]. The lower the tolerance, the more likely is multicollinearity to exist among the variables. The independent variables are not correlated if the VIF equals to 1. However, a VIF value between 1 and 5 implies that the variables are moderately correlated. The challenging case is when the VIF value is between 5 and 10, as it indicates highly correlated variables. If VIF ≥ 5 to 10, then multi-collinearity exists among the predictors in the model. VIF > 10 implies that the regression coefficients are weakly predictable in the presence of multi-collinearity [21].

Dummy variables

When nominal independent variables such as sex and race, or ordinal variables with unequal distance between levels of the category, such as degree of severity (low, medium, and high), are entered into the regression model, the investigator is required to transform an independent variable into a set of dummy variables. For an independent variable with level "a," "a-1" dummy variables can be defined [4, 6]. For example, for a variable with three severity levels (low, medium, and high), two dummy variables (D_1 and D_2) are created as follows:

$$D_1 = \begin{cases} 1 & \text{if severity is low} \\ 0 & \text{otherwise} \end{cases} \qquad D_2 = \begin{cases} 1 & \text{if severity is medium} \\ 0 & \text{otherwise} \end{cases}$$

The "high" severity level is considered as a reference or the baseline category [6]. The reference category should be selected based on the expert opinion of the clinical investigator.

Sample size in regression

For sample-size calculation with power analysis, we are required to know the desirable power, the results of one- or two-sided tests, the common variance, the slope for detection, and the number of independent variables [22, 23]. The greater the number of independent variables, the larger is the sample size. Generally, 10 or even up to 20 participants are recruited for every independent variable [5, 10].

Common mistakes in reporting linear-regression analysis

The common mistakes in reporting a regression model are as follows:

1. Several assumptions must be verified before considering a linear-regression model [5]; however, they are typically not verified. If these assumptions are not satisfied, then a nonlinear regression or other statistical analyses should be performed.
2. However, the model may not be compatible with several variables. When a dependent variable and multiple independent variables are included in a linear-regression model, multiple regression is appropriate but not multivariate regression [4, 10]. This is because the multivariate model contains more than one dependent variable [10].
3. Standardized and unstandardized regression coefficients are recommended to be reported together [17, 18]; however, this is not presented in some publications.
4. Variable transformations are not used sensibly [10]. When independent variables are nominal or ordinal with unequal distances between categories, the dummy variables should be defined and entered into a linear-regression model [4, 6].
5. An expert statistician is not assigned from the development of the research idea until the final revision of the manuscript is published. Appointing an expert statistician as an integrated team member is a judicious approach for using mathematical principles optimally and offers scholarly insights into the benefits of science development [24, 25].

Recommended checklist for editors and peer reviewers

The following is a comprehensive checklist to ensure that the linear-regression model is conducted appropriately and yields valid results:

1. Clarity of the title, aim, and research questions

A well-formulated research question, title, and aim of the study can guide the selection of variables and the interpretation of outcomes [10].

2. Data acquisition

Verify the definition of dependent and independent variables while verifying for missing values, outliers, and inconsistencies [2-7]. Ensure that data are obtained systematically to minimize potential bias. In the presence of a nominal or ordinal independent variable, determine the appropriate transformation [4, 6]. Verify the appropriateness of the sample size based on the number of independent variables included [5, 10, 22, 23].

3. Model selection

The simple linear regression model should be chosenif the relationship between a single dependent and independent variable is tested. However, multiple linear regression is suitable if a single dependent variable and multiple independent variables exist to investigate the relationship between these variables in combination to the dependent variable [3, 5, 6, 11, 14].

4. Assumptions

Linear regression relies on several assumptions, including normality, equality of variance, linearity, and collinearity, which must be satisfied before proceeding with the analysis [4, 10, 19, 20].

5. Reporting and interpretation

The report of the findings should be clear and comprehensive, and the tables, plots, and diagrams must be consistent. Standardized and unstandardized regression coefficients should be reported simultaneously [17, 18] at a relevant significance level [10]. When additional independent variables are added to the model, the R^2_{adj} must be reported [3, 10]. The interpretation of the results should be consistent with the context of the research question. The first step is to verify the significance of the correlation and then consider the R^2 for the strength of this significant correlation. Verify if the results are overinterpreted, particularly in correlational contexts, as the results should be based on the R^2 [10].

CONCLUSIONS

The linear-regression model is a commonly used statistical method in research, including research in vision sciences. This method can provide valuable insights when conducted with rigor and attention to detail. The appropriate selection of the regression model and the presence of model variables are key measures that must be established and controlled strictly to achieve valid statistical results. The results may be inaccurate if an appropriate regression model is not established.

ETHICAL DECLARATIONS

Ethical approval: Not applicable.

Conflict of interest: None.

FUNDING

This review article has been funded by IVORC Academic Foundation, Texas, United States.

ACKNOWLEDGMENTS

None.

REFERENCES

1. Barros AJ, Hirakata VN. Alternatives for logistic regression in cross-sectional studies: an empirical comparison of models that directly estimate the prevalence ratio. BMC Med Res Methodol. 2003 Oct 20;3:21. doi: 10.1186/1471-2288-3-21. PMID: 14567763; PMCID: PMC521200.
2. Stephenson J, Bunce C, Doré CJ, Freemantle N; Ophthalmic Statistics Group. Ophthalmic statistics note 11: logistic regression. Br J Ophthalmol. 2016 Dec;100(12):1594-1595. doi: 10.1136/bjophthalmol-2016-309223. Epub 2016 Nov 3. PMID: 27811279.
3. Darlington RB, Hayes AF (2016). 'Regression analysis and linear models: Concepts, applications, and implementation'. (pp: 1-155). Guilford Publications. ISBN 9781462521135.
4. Montgomery DC, Peck EA, Vining GG (2021). 'Introduction to linear regression analysis'. (pp: 1-114). John Wiley & Sons. ISBN: 978-1-119-57872-7.
5. Zapf A, Wiessner C, König IR. Regression Analyses and Their Particularities in Observational Studies. Dtsch Arztebl Int. 2024 Feb 23;121(4):128-134. doi: 10.3238/arztebl.m2023.0278. PMID: 38231741; PMCID: PMC11019761.
6. Kasza J, Wolfe R. Interpretation of commonly used statistical regression models. Respirology. 2014 Jan;19(1):14-21. doi: 10.1111/resp.12221. PMID: 24372634.
7. Iqbal M, Elmassry A, Elgharieb M, Said O, Saeed A, Ibrahim T, Kotb A, Abdelhalim M, Shoughy S, Elgazzar A, Shamselden H, Hammour A, Eid M, Elzembely H, Abdelaziz K. Visual, ocular surface, and extraocular diagnostic criteria for determining the prevalence of computer vision syndrome: a cross-sectional smart-survey-based study. Med Hypothesis Discov Innov Ophthalmol. 2024 Jul 1;13(1):1-15. doi: 10.51329/mehdiophthal1489. PMID: 38978825; PMCID: PMC11227667.
8. Lotfy NM, Shafik HM, Nassief M. Risk factor assessment of digital eye strain during the COVID-19 pandemic: a cross-sectional survey. Med Hypothesis Discov Innov Ophthalmol. 2022 Dec 3;11(3):119-128. doi: 10.51329/mehdiophthal1455. PMID: 37641641; PMCID: PMC10445314.
9. Keeling KB, Pavur RJ. A comparative study of the reliability of nine statistical software packages. Computational Statistics & Data Analysis. 2007 May 1;51(8):3811-31. doi: 10.1016/j.csda.2006.02.013.
10. Baždarić K, Šverko D, Salarić I, Martinović A, Lucijanić M. The ABC of linear regression analysis: What every author and editor should know. European science editing. 2021;47:1-9. e63780. doi: 10.3897/ese.2021.e63780.
11. Bali J, Bali O. Artificial intelligence in ophthalmology and healthcare: An updated review of the techniques in use. Indian J Ophthalmol. 2021 Jan;69(1):8-13. doi: 10.4103/ijo.IJO_1848_19. PMID: 33323564; PMCID: PMC7926114.
12. Savini G, Hoffer KJ, Barboni P, Schiano Lomoriello D, Ducoli P. Corneal Asphericity and IOL Power Calculation in Eyes With Aspherical IOLs. J Refract Surg. 2017 Jul 1;33(7):476-481. doi: 10.3928/1081597X-20170504-05. PMID: 28681907.
13. Savini G, Hoffer KJ, Barboni P. Influence of corneal asphericity on the refractive outcome of intraocular lens implantation in cataract surgery. J Cataract Refract Surg. 2015 Apr;41(4):785-9. doi: 10.1016/j.jcrs.2014.07.035. PMID: 25840302.
14. Maulud D, Abdulazeez AM. A review on linear regression comprehensive in machine learning. Journal of Applied Science and Technology Trends. 2020 Dec 31;1(2):140-7. doi: 10.38094/jastt1457.
15. Cha DS, Moshirfar M, Herron MS, Santos JM, Hoopes PC. Prediction of Posterior-to-Anterior Corneal Curvature Radii Ratio in Myopic Patients after LASIK, SMILE, and PRK Using Multivariate Regression Analysis. J Clin Med. 2023 Jul 7;12(13):4536. doi: 10.3390/jcm12134536. PMID: 37445571; PMCID: PMC10342661.

16. Zareei A, Razeghinejad MR, Salouti R. Corneal Biomechanical Properties and Thickness in Primary Congenital Glaucoma and Normal Eyes: A Comparative Study. Med Hypothesis Discov Innov Ophthalmol. 2018 Summer;7(2):68-72. PMID: 30250855; PMCID: PMC6146241.
17. Abdelshafy M, Abdelshafy A. Correlations Between Optical Coherence Tomography Angiography Parameters and the Visual Acuity in Patients with Diabetic Retinopathy. Clin Ophthalmol. 2020 Apr 23;14:1107-1115. doi: 10.2147/OPTH.S248881. PMID: 32425497; PMCID: PMC7186882.
18. Grace JB, Bollen KA. Interpreting the results from multiple regression and structural equation models. Bulletin of the Ecological Society of America. 2005 Oct 1;86(4):283-95. doi: 10.1890/0012-9623(2005)86[283:ITRFMR]2.0.CO;2.
19. Mertler CA, Vannatta RA, LaVenia KN (2021). 'Advanced and multivariate statistical methods: Practical application and interpretation'. (pp: 176-210). Routledge. doi: 10.4324/9781003047223.
20. Schmidt AF, Finan C. Linear regression and the normality assumption. J Clin Epidemiol. 2018 Jun;98:146-151. doi: 10.1016/j.jclinepi.2017.12.006. Epub 2017 Dec 16. PMID: 29258908.
21. Shrestha N. Detecting multicollinearity in regression analysis. American Journal of Applied Mathematics and Statistics. 2020 Jun 16;8(2):39-42. doi:10.12691/ajams-8-2-1.
22. Jenkins DG, Quintana-Ascencio PF. A solution to minimum sample size for regressions. PLoS One. 2020 Feb 21;15(2):e0229345. doi: 10.1371/journal.pone.0229345. PMID: 32084211; PMCID: PMC7034864.
23. Serdar CC, Cihan M, Yücel D, Serdar MA. Sample size, power and effect size revisited: simplified and practical approaches in pre-clinical, clinical and laboratory studies. Biochem Med (Zagreb). 2021 Feb 15;31(1):010502. doi: 10.11613/BM.2021.010502. Epub 2020 Dec 15. PMID: 33380887; PMCID: PMC7745163.
24. Heidary F, Gharebaghi R. COVID-19 impact on research and publication ethics. Med Hypothesis Discov Innov Ophthalmol. 2021 May 31;10(1):1-4. doi: 10.51329/mehdiophthal1414. PMID: 37641621; PMCID: PMC10460218.
25. Brownstein NC, Louis TA, O'Hagan A, Pendergast J. The Role of Expert Judgment in Statistical Inference and Evidence-Based Decision-Making. Am Stat. 2019 Mar 20;73(1):56-68. doi: 10.1080/00031305.2018.1529623. PMID: 31057338; PMCID: PMC6474725.